DOCUMENT RESUME

ED 190 097 IR 008 579

AUTHOR Hilton, Howard J.
TITLE HUC--A User Designed System for All Recorded Knowledge and Information.
PUB DATE Mar 80
NOTE 23p.: Paper presented at the National Online Information Meeting (New York, NY, March 25-27, 1980). Not available in paper copy due to marginal legibility of the original.

EDRS PRICE MF01 Plus Postage. PC Not Available from EDRS.
DESCRIPTORS *Information Retrieval: *Information Systems: *Online Systems: Research Projects: Research Proposals: *Systems Development: *Technological Advancement
IDENTIFIERS Hilton United Corporation: Videotext

ABSTRACT

This paper proposes a user designed system, HUC, intended to provide a single index and retrieval system covering all recorded knowledge and information capable of being retrieved from all modes of storage, from manual to the most sophisticated retrieval system. The concept integrates terminal hardware, software, and database structure to allow any literate person, without special training, to ultimately obtain access to all available public information stored in any of the multitude of online and videotext systems. It would also use broadcast videotext for current information: topical and index updates: and, for cable systems, TV programs could also be stored during the night on videodiscs. HUC is so structured that as voice command systems develop, the user can use the system without a keyboard. The paper details the assumptions and principles of the HUC concept: a unique identification system: index structure: unified software: terminal hardware: mass density storage: system configuration: and the benefits and applications of the system. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED190097

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

BEST COPY AVAILABLE

BEST AVAILABLE COPY

HUC - A USER DESIGNED SYSTEM FOR ALL RECORDED KNOWLEDGE AND INFORMATION

by

Howard J. Hilton

BEST AVAILABLE COPY

BEST AVAILABLE COPY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Howard J. Hilton

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

IR008519

HUC - A User Designed System for All Recorded Knowledge and Information

by

Howard J. Hilton

Paper prepared for the session, "Information Systems of the Future" of the National Online Information Meeting New York, March 25-27, 1980

Copyright 1980 by HUC (Hilton United Corporation)
Printed electronically in the United States of America
by CompuServe Incorporated

I. INTRODUCTION

This is a summary of one portion of a longer work entitled "Videotext, Computers, and HUC -- Past, Present, and Future". This work is conceived as a review and prognostication which would have sections of interest to professionals, students, and those concerned with developments which may have a profound effect on their lives. Since it is designed for online retrieval and on demand reproduction, the completed work will not be subject to the market limitations nor space constraints of a book. Hopefully it will introduce the new era in which an individual can scan a work and select for reproduction in a library, a store offering such service, or at home, or at work, only those sections of interest.

Think of the advantage of such a system for marketing, or for professors assigning reading for students, the research ease for libraries, and many other situations. An individual would be able to enter a work from a terminal at home and send it to an editor by telephone, cable TV or other means, or make it immediately available worldwide where it can be brought to the attention of those interested by a computer programmed to scan on a daily basis the latest index entries for the selective dissemination of information. This is just one aspect of the current process of change.

The history of human civilization is marked by fundamental changes which have occurred in the social, economic, and political structure. The first is measured in thousands of years in the transition from nomadic to agricultural. The next from

agricultural to industrial can be measured in centuries. The present transition from an industrial society to one that has been variously termed post industrial society, the information society, the electronic society, the technological society, etc. can be measured in decades.

As with any fundamental social change, the implications of which are dimly perceived by those who are deeply involved in it, the political and economic reverberations often appear as the dominant elements, when in reality they are frequently the result of initiatives of individuals or groups taking advantage of the forces of change or reacting to them.

The potential for good or evil of this process of social change, which began this century, ranges from the realization of utopian dreams to the intellectual and physical enslavement of people and nations or even the destruction of life on this planet. If one of the utopian dreams, the democracy of the intellect, so elequently described by Bronowski, is to be realized, one of the results of this transition will be that before the end of this century, nearly every family and office and library that now has a telephone and TV or can afford them will have access to virtually all public knowledge and information available in digital form. Individuals will be able to maintain better records, in many cases to work at home, to make their work available to anyone interested, and to be assured of protection and privacy for personal records.

The development of very large scale integration, making possible faster, more efficient, and less expensive devices for distributing information, is well known, as are the various developments for faster, denser, and less expensive memory. This technology is making possible the introduction and marketing of various videotext systems. Videotext is divided into two types, broadcast and interactive. Videotext is a term that has gained currency with the development of systems using a decoder with a TV set for receiving and displaying alphanumeric and graphic information. Thus, it would include existing systems such as viewdata, CEEFAX and Oracle, Antiope, Telidon, Captain and similar systems.

With all the different systems, containing all types of information available to the public, the question is --- How can the individuals in the home or office obtain information concerning what is available and then obtain access to it? This presents both technical and institutional problems. One technical problem is the interface of different types of systems; the other is the organization of all the data in all of the different systems, access to it, and payment for it, which are basically institutional problems.

It is clear that the developments in technology and the introduction of videotext systems will require new terminal hardware capable of handling both videotext and current online systems. If this is the case, then the terminal should also be

designed to provide the most efficient retrieval of information from the multitude of sources. This paper describes such a system.

The HUC concept integrates the terminal hardware, software, and database structure as a system that any literate person without special training may use, ultimately, to obtain access to all available public information stored in any of the multitude of online or videotext systems. It would also use broadcast videotext for current information, topical and index updates, and for cable systems, TV programs which could be stored during the night on videodisks. HUC is so structured that as voice command systems develop, the user can use the system without a keyboard. Its objective is to provide a single index and retrieval system covering all recorded knowledge and information capable of being retrieved from all modes of storage from manual to the most sophisticated retrieval system.

The remaining sections of this paper, in addition to the Introduction, are ---

II. Assumptions and Principles of the HUC Concept
III. Unique Identification
IV. Index Structure
V. Unified Software
VI. Terminal Hardware
VII. Mass Density Storage
VIII. System Configuration
IX. Benefits and Applications
X. Conclusions

II. ASSUMPTIONS AND PRINCIPLES OF THE HUC CONCEPT

The HUC system concept is based upon the following principles and assumptions:

1. An effective informations system requires the unique identification, immediately upon entry, of all material to be stored and retrieved.
2. The user requires more than a digital code as an interface in accessing large databases.
3. The effective retrieval of information requires extensive indexing and an algorithm which helps sift the wheat from the chaff and the new from the old entries.
4. A system for all recorded knowledge and information requires a unified software which (a) will facilitate retrieval in all media, electronic, mechanical, and manual, (b) will protect information not publicly available, (c) will account for payment where required, (d) will be equally effective for new and sophisticated users on personal or large computer systems, and (e) will provide a permanent storage, which cannot be eliminated or altered without the destruction of the mass storage medium.

5. Users should be able to use the system for electronic communication and for the storage and retrieval of their private files using the same software as for public information.
6. For electronic retrieval and communication, terminals should range in cost and benefit from keypads using standard TV sets to multipurpose, multilingual keyboards with high resolution color CRTs.
7. Terminals to be used for communication, maintenance of personal or organizational records, or for electronic transfer of funds and accounting for payments should have special encryption devices.
8. Using existing and developing resources including libraries, library networks, teletext, viewdata, and online systems now in place, cable TV, FM subcarriers, satellite networks, videodisks and other mass density storage media, such a system as HUC is now both feasible and economically beneficial for the users, producers, and purveyors of knowledge and information.

Given the assumption that a system requires unique identification of all material in it and that a user needs more that a digital code for retrieval, the problem in designing a system for all recorded knowledge and information is to reconcile the different requirements of computer and users.

Computers require the shortest unique binary identification for all material in the system.

Users require the optimum amount of information to make an initial selection of desired material. At present this is usually the library catalog card, an abstract, or a citation.

The efficient operation of a total system of the magnitude under consideration here requires something in between which can be used as an interface between the user and the computer, so that the requirements of both are met. The first essential element is the unique identification in a structured format, which is important for computer manipulation. The other essential elements which utilize this unique identification are indexes, unified software, terminal hardware, encrpytion devices, mass density storage, and both terminal and personal addresses.

These essential elements are combined and supplemented by electronic communications and transfer of funds in a system configured to meet the following objectives:

1. Providing the means by which all recorded knowledge and information can be stored, retrieved, and distributed,

2. Enabling users to communicate and to obtain knowledge and information promptly and in the medium of their choice, optimizing cost, convenience, and other factors,

3. Assuring authors and producers that they can prsent their work to the world at the least cost and that they will receive fair compensation for their intellectual effort, and

4. Protecting the privacy of non public records and communications.

Both government and private research funds have been utilized in creating the technology for achieving these objectives. In contrast little public effort has been devoted in the United States to the solution of the technical problems of retrieval, and the institutional problems involved in storing, retrieving, and distributing the world's knowledge. The HUC system is a concept of database organization and distribution, software, hardware, banking, and online systems that would utilize existing technology to provide more effective and equal access to the world' knowledge.

III. UNIQUE IDENTIFICATION

The universal code for the unique identification of all recorded knowledge and information (HUC) is designed to be (1) the interface between the user and the computer, (2) the foundation for unified software, (3) the basis for extensive indexing, and (4) the basis for electronic, mechanical, and manual retrieval and for security, payment, and communication.

The identification of all material by this code permits a simplified and unified software to provide access to documents, records, and all types of material in all different media. The basic HUC storage medium is either mass density optical disk with a reduction ratio of 450 to 1, or high reduction microforms with a reduction of 200 to 1, depending upon the requirements of the users. Where microforms are used, whether high reduction or not, each frame has the unique HUC identification. This makes possible the recovery of full text of documents from which only a copy of a page or section may be available.

The code utilizes the HUC character set. This character set extends the alpha characters to 32 and adds 31 using a reverse slash through the capital letters and the additional characters. In addition, it designates 26 numeric characters. The reverse slash in the alpha characters serves several functions. In the geographic segment identification, it indicates the quadrants. It also provides a one character identifier for time or geographic measurement in minutes and seconds.

The code is mnemonic, alternating alphabetic and numeric fields. The total number of characters is 61, the highest prime number that can be contained in two packets of 32 bytes each of a packet switching network. There are a total of fifteen fields of which three are for check characters designed to confirm the accuracy of entry and transmission of the preceding fields of the code. Normally, unique identification would be provided in less

than one half of the fields and about one third of the number of characters. The fields of the code and the number of characters in each are indicated below:

1. Type of material -- 2 numeric, or for official national or international publications or documents, or for statistical series, 2 alpha characters;
2. Source -- 6 alpha and 5 numeric characters;
3. Check character -- 1 alpha character;
4. Status of material -- 1 numeric character;
5. Time by year, month, day, hour, minute, and second, or edition or number of a document or work within any of these periods from 1700 to 2700 -- 7 alpha characters;
6. Primary storage identification --2 numeric characters;
7. Check character -- 1 alpha character;
8 Special media designation -- 1 numeric character;
9. Special media identification -- 3 alpha characters;
10. If microform, type or other media designation -- 1 numeric character;
11. Publisher, producer, copyright royalty account, or author in special situations -- 6 alpha and 5 numeric characters;
12. Language fields -- 2 alpha character;
13. Special field designation -- 2 numeric characters;
14. Special field classification, document numbers, mass density storage addresses etc. -- 2 alpha and 13 numeric character; and
15. Check character -- 1 alpha character.

The mnemonic features for periodical literature are based upon the CODEN. For individuals the first three letters of the last name are used with the remaining three alpha and five numeric characters to maintain alphabetical order, which is important when the database is to be used for manual searching. Countries within regions are identified by appropriate letters. In addition to the source of material, the other substantive fields indicate the legal status, date to the second if necessary, electronic availability, other media identification, copyright owner or author if official government document with author identification, languages, and two fields which serve a multitude of purposes such as the subject classifications of the Library of Congress, Universal Decimal Classification, addresses in mass memory storage, spare part numbers, uniform product codes, document numbers, etc..

The only centralization required for the initial operation of the system is the assignment of source codes to countries, organizations and multinational corporations and blocks of digital equivalents for the daily files. Organizations creating knowledge and information and using the system for communication, transfer of funds, or for maintaining private and confidential records can maintain the system on a decentralized basis once these assignments are made. With a source identification, the creator can uniquely identify all information entered in the

system by time and other designation, if necessary.

IV. INDEXES

The unique identification also serves as the basis for extensive indexing. The HUC index is divided into five categories. Documents can be indexed one or more times in three categories: subject terms, proper names, and geographic. They are entered only once under the designations for type and level of material. The index terms are also identified by the code which permits the identification of three million or more subject terms or phrases, proper names in the billions, geographic names in the millions and geographic segments in the trillions. The geographic segment index is based upon multiples and divisions of four degree by six degree segments. These segments require only three alpha characters for identification with the HUC reverse slash characters indicating the quadrant of the earth. The one second segment requires ten alpha numeric characters.

The HUC identification also permits the identification and provides a structure for indexing information for all of the stars in the billions of galaxies. Worldwide addresses for individuals, organizations, terminals, and sensing devices, with time in nanoseconds for a thousand year range are also provided by the code and indexing structure.

Indexes for the first three categories are subdivided by type of index. The subject category, for example, would identify thesaurus controlled terms, library subject headings, computer indexed abstracts or, in the most sophisticated cases the syntactic analysis of full text by special computer program. The proper name index likewise has a number of subdivisions ranging from telephone numbers through citations by and to, legal decisions, and biographic statements without the space limitations of "Who's Who." The geographic index is by names and places, administrative or physical areas, geographic segments, and by other methods of identifying position.

The index terms reference the complete document and the index unit. The index unit depends upon the structure of the original document; if a book published by traditional methods, it would in all probability be a page. In the new electronic media, it could be a frame, a single statistical element, or a group of such elements. It could also be the individual frame of a TV program, or satellite images. In other words, the index unit is the smallest element to which reference might be made.

V. UNIFIED SOFTWARE

The unified software for providing access to recorded knowledge and information has ten major elements: (1) organization, (2) identification, (3) user guidance and assistance, (4) selection (5) access, (6) manipulation, (7) accounting and payment, (8) entry, (9) communication, and (10) control security.

ORGANIZATION: The organization of material in the system is based upon five files: (1) the source file, which includes all code assignments for the first two fields of identification code, (2) the daily document file, (3) index file, (4) format file, and (5) file of chain relationships. Each of these files is identified by one of the six hexadecimal numeric characters of the HUC character set. Each file contains the computer numeric code equivalent for the unique identification code. These files are the central feature of the unified software enabling it to organize the access and stand alone software for the entire system dealing with the multitude of media, equipment, and requirements.

IDENTIFICATION: The software for identification reformats and identifies material, assigns index terms, creates document profiles, and enters the permanent storage address and the final check characters.

USER GUIDANCE AND ASSISTANCE: Most systems today require the reference to volumes of manuals by the user for effective use of the system. The object of unified software is to provide the user with answers within the system to questions now answered in the manuals. By use of default modes and queries, a literate user who had never touched a computer terminal should be able to use the system immediately as one now uses a catalog in a library, reference books, a table of cases, "Who's Who", parts lists, repair manuals, telephone directories, statistical tables, citation indexes, sales catalogs, announcements, brochures, and other information. This software is based on five modes of operation: command, program, retrieve, compose, and communicate.

SELECTION: Today for the individual, selection of knowledge and information for entertainment, edification, or answers to specific questions is severely limited by the random offerings of the news and entertainment industry, the author and subject classifications of libraries and index services, and many other information programs and enterprises. Unfortunately, the identification of the available possibilities is so proliferated or controlled that the individual is at the mercy of the system instead of the system being the servant to the individual.

By identifying and indexing all available databases and other material in all media, the HUC system enables the user to make more precise selections. Since the magnitude for some index terms is overwhelming, especially for individuals with limited search capability, a central feature of the selection software is an algorithm designed to identify the most important items in the first three categories of the index. The principal elements involved in this selection are time, use, evaluation, type, and level of material.

For more definitive searches in more powerful systems, the selection process is based upon the program statement which contains five fields in 80 columns identified and described as follows:

4 numeric characters for sequence or command identification,
4 alpha characters for commands,
5 numeric characters to identify fields of operation,
2 alpha characters for logical operations,
65 alpha or numeric characters describing files and programming for data file manipulation.

ACCESS: In many systems offering a variety of databases today, it is necessary to use a different search strategy for each database. The use of terms that produces an acceptable number of documents, and the user either looks at abstracts or receives a bibliographic description. In the HUC system concept, one could search the entire database with a single strategy or each of the different categories and types of databases separately with the strategies designed for the database desired. Where appropriate arrangements would exist, the user would be transfered automatically to the online system with the database selected. In either case the user has an interface with the HUC mnemonic code to assist in getting a feel for the volume and diversity of the literature or available data.

The next step is to look at the document profiles, which would contain available bibliographic material as well as all index terms and descriptions, or at the abstracts, if available. The object is, of course, to produce the document, which, as used here, includes any work in any medium that could be covered by copyright or is publicly available. If the document is in the mass density storage in the system, it can be produced by a single command. If it should be a TV program, it would probably be loaded into a videodisk recorder during the night. The HUC identifier would provide the signal to activate the recorder. If it is not in the system, the document profile or abstract should provide the means for ordering a copy. This software would also provide a list of documents relevant to one previously entered in the system.

MANIPULATION: The unified software provides for overriding the program by commands for specific modes or for calling other programs into the terminals or for operation in an online mode. The software would provide for the manipulation of data already available as documents in the system by reference to the HUC format identification and appropriate rows and columns. The statistical data files, which are not organized in tables but available for manipulation by the user, are divided into four categories: (1) correlative subjects, (2) subject variables, (3) geographic variables, and (4) time variables. Each of the data elements can be associated with the appropriate field and identifier of the HUC identification. Thus various sources could develop data by census tracts, zip codes, or other geographic, subject, or time categories.

Bibliographic files can be similarly manipulated with Boolean or full text search techniques. As indicated in the discussion of selection, the software provides for the manipulation of the code for eliminating documents on the basis of time, language, medium, source, and legal status and for selecting only those that meet certain requirements such as classified subject heading or subject categories identified in various catalog systems.

ACCOUNTING AND PAYMENT: While every person should have effective access to copyrighted material, creators of intellectual property should have fair compensation for their effort. The HUC concept provides for compensation for use of copyrighted and proprietary material on a standard scale or at special rates set by the copyright owner. The standard scale would provide for a low payment for the material in the mass density store which would be widely distributed to libraries SDI (Special Dissemination of Information) centers, and the accounting and switching centers. They in turn would pay for all reproduction or dissemination to clients beyond "fair use". Libraries would not pay for use of material on the premises; however, a charge would be made for copies similar to the charge the user pays for photocopying in the library. The difference is that a portion of the charge would be for copyright royalty or database subsidy in the case of material in the public domain. The payment, which would be lower for microform reproduction than for hard copy, would be made to the account designated by the code.

The payment system would also handle orders placed through the system for goods and services and for billing for electronic communication. The day may well come when packaged food will be ordered from a central warehouse, automatically collected from bulk containers, delivered to the home by truck and paid by electronic transfer of funds. Think of the saving in gasoline when a thousand cars at the supermarket are replaced by a couple of trucks delivering groceries, just like the good old days.

ENTRY: The software for entry consists of identification and address, word processing, restricted or worldwide access, indexing, and special handling if required. The system would provide for circulating manuscripts and maintaining privacy until the delivery of a paper or publication. At that time it could be entered in the system, copyrighted in the name of the author or publisher, and released for worldwide dissemination in the mass storage medium, with payments for reproduction to be received and deposited to the owner's account at a standard or special rate.

COMMUNICATION: Electronic communication in the HUC concept could be sent to an individual, an organization, or a terminal or to any combination. The recipient of a communication would be distinguished from other material by a reverse slash character at the beginning of the address code. When entered in the system, the time to the second would be entered in the code. The sender would be identified by a reverse slash in the second column.

Slashes in both columns would indicate a completed exchange.

CONTROL AND SECURITY: One of the problems of computerized systems is the difficulty of maintaining and audit trail in a medium that can be so easily altered. One of the principles of the HUC system is that any public document or required record should be preserved in a mass density store which, once written cannot be changed. The file of chain relationships would provide, however, that any requests for a document would also produce corrections and subsequent changes as well as comments.

With a mass density store in the order of gigabits per square inch, and relative low cost for reproduction, the cost of keeping such records would only be a fraction of the present cost for paper records, not to speak of the space.

Security is provided by the initial indication in the field of the code for legal status. If material is private or confidential, encrpyption would normally be employed. In addition, the present security in the form of password and counterchecks of individuals and terminals would be used and enhanced with the authentication capability of the encryption device. The highest level of security is provided by the mass density store with encrypted files, delivered by hand, or created and used in a stand alone system in a secure area.

The unified software would operate at various levels:

1. a central level for assigning source codes and blocks of four bit hexadecimal computer codes for the daily files,
2. the regional or system nodes,
3. the international online and communications networks,
4. local distributive and accounting networks, and
5. stand alone system and HUC terminals.

Without a system of universal and unique identification, the benefits of much of the new technology will be denied to the public at large. With it we can develop systems that will utilize all media and make it possible for the consumer to buy a product, take it home, and begin to enjoy access to all the knowledge, information, and entertainment that society has produced and will produce in the future.

VI. TERMINAL HARDWARE

Although the HUC database can be easily accessed in public viewdata, teletext, cable TV, and similar systems by a keypad of twelve characters or a standard keyboard, the HUC keyboard or terminal would make such systems much more effective. The HUC keyboard would differ from the standard ones, since in addition to other modifications, it would automatically display the HUC character set when the codes are displayed. The shift is identified by the use of the escape sequence for a private code. It is activated by the HUC key or by the appropriate designation

in the identification software

For the multilingual keyboard, the HUC identifier would be used to shift to the appropriate characters used in the French, German and other languages, the characters for which would be contained in the HUC terminal or keyboard.

All HUC terminals would contain a Cryptext unit which is an encryption device developed by the Cryptext Corporation. This device would contain the HUC terminal identification which would also be the key for terminal authentication and for controlling the use of software and other proprietary informatio The terminal would also contain a multilingual character generator. Although only more sophisticated terminals could utilize the full capability of the HUC chips, in the volume of production required for the future market, it would probably be cheaper to use the same chips for all key pads and terminals with the possible exception of the encryption capability.

The possibilities at the present would be the pad of 12 keys of the viewdata system, the 16 key pad as proposed for the "Green Thumb" project, the standard type keyboard with the HUC key and shift controls which could be the first level of the keyboard terminal. Additional levels would add cursor controls, viewdata and other graphic controls, numeric block, communications controls, peripheral keys, supplementary language keys, and special function keys.

The total system software is developed around the concept of a terminal keyboard with a memory buffer and disk or tape for additional storage. In addition it would contain (a) the HUC chip with the character generator for the HUC character set, all characters of the Roman alphabet, the Kata Kana characters of the Japanese Industrial Standard Code for Information Exchange, and the Cyrillic alphabet, (b) terminal identification program, (c) encryption device, and (d) fiber optic connector with RF modulator if to be connected to a TV set.

VII. MASS DENSITY STORE

Some writers have predicted the development of an optical disk capable of providing random access to 10/15 bits. Since for a 12 inch disk, this would be an order of magnitude beyond the present theoretical capability of electron beam recording, the mass density store should probably be predicated upon a density of 10/8 or 10/9 bits per square inch. According to the Department of the Navy, "...technology is available to provide a mass store with access of two megabits per second and a capacity of 2 x 10/12 bits within a 2 1/2 foot cube." (Report No. MI-028, February 1976)

In any case there is considerable work in the area of mass density storage media. These include magnetic media of increased density, holographic fiche, slides, and tapes, solid state devices, optical and electron beam recording and laser devices.

Although video recording units commercially available are capable of capturing the neighborhood of 10/10 bits, the error rate is presently too high for this to be used as the mass density store.

A paper presented by Jerome Drexler at the ELECTRO-OPTICS/LASER '79 conference indicates that optical disks are now available with a media cost of $0.10 per megabyte, a storage capacity of 900 megabytes for one side of a twelve inch disk, a data recording rate of ten megabits per second, and a reading rate of more than twice that. The raw error rate of the laser recording media which his company has developed is reported to be 1 in 10 megabits.

VIII. SYSTEM CONFIGURATION

Until the development of radio, access to knowledge and information was provided by what we now call hard copy. As we move further into the electronic age, hard copy may be produced in the home by electronic and photo-optical techniques or delivered in response to an electronic request: for savings in cost and resources, it may be a microform copy. In any case, the heart of the system for access to knowledge and information will be the terminal.

Under present arrangements, an individual with a terminal in the home can access a variety of databases on online systems, but no single place can provide a total index. Indeed, searching a variety of databases requires entry into a number of systems using different search techniques and strategies. The configuration under the HUC concept, which would provide a total index, would involve (1) a central level for the initial organization of the database, (2) the regional or systems nodes for the compilation of indexes, basic store of material, distribution, and initial communications switching, (3) international online and communications networks, (4) local distribution and accounting networks, and (5) stand alone systems, terminals, and intelligent keyboards.

CENTRAL LEVEL The HUC system can be developed with a strong or weak central element. With a strong central element. Initial source codes would be issued and maintained by a single unit connected by open circuit to the regional or system nodes. The database organization likewise would be compiled, maintained, and distributed in the mass density store or electronically to the regional or system nodes.

With a weak central element, the source codes would be distributed in blocks to the regional or systems nodes which would maintain the detailed records. The initial assignment could be in the form of an agreement among the appropriate authorities of the various countries participating in the system.

The present environment would seem to favor private initiative. A nonprofit institution could very well undertake the central function as an adjunct to an international

communication or an online systems network. Other possibilities include an existing publishing operation, a hardware development and production concern, an information provider on public viewdata type systems, a cable TV operator, or a banking and credit card institution with international interests.

If this course were to be followed, the HUC organizer, on the basis of appropriate agreements, would take (1) all available databases which can be publicly accessed in the United States, through EURONET, viewdata and teletext systems, and other digital delivery systems, (2) newspaper, periodical, and other material which is computer generated, and (3) the system input of individual creators. It would then assign HUC identifiers to all documents and would index them for worldwide distribution on the basis of agreements concerning the reproduction and payment of copyright royalties. The index would also be available at all levels.

Since libraries have traditionally been the means of providing access to knowledge and information, they and their networks are central to any program for the future. It is the objective of the HUC system to mobilize their resources at all levels. This would involve the cooperation of the several library networks which are in operation in various countries today. In consolidating their bibliographic databases in the multilingual index, they and all other users of the system will benefit. It should also facilitate the gradual conversion of existing holdings to a digital database, since the index and HUC identification create a structure applicable to all types of existing systems.

REGIONAL OR SYSTEM NODES The regional or system nodes at the present time are primarily online networks and public viewdata type systems. They can play an important role in providing access especially during the transition period when digital databases are covering a broader spectrum of knowledge and information and are becoming more generally available to the public. If they are part of the HUC system, their primary role would be that of providing the index and access to other databases. In doing this they would presumably use the algorithm for selecting about 250 items per term which would most likely meet the requirements of their subscribers.

The software for the regional or system nodes is basically the database management software of the online systems or the tree structure storage, retrieval, and accounting software of viewdata systems. The additional software for the HUC system would be designed to make the terminals and the system more effective. It would do this by taking index and substantive dumps, providing off line searching, and dumping requests for action. In addition the software would create the window by which HUC terminals could enter the system from local distributive or accounting networks or go through the system to international and communication networks.

INTERNATIONAL ONLINE SYSTEMS AND COMMUNICATION NETWORKS The indexes would contain the addresses for abstracts, reviews, and documents for works and data contained in the various bibliographic and statistical databases available on international online networks. It would also contain the addresses of terminals, organizations, and individuals for use by international communication networks. Thus a subscriber using the HUC identification code should be able on a viewdata system to go through the HUC window to contact

an international network either for information or communication.

Since payment has to be made for such use of the system, the accounting has to be performed at one or several levels. At present under EURONET, the subscriber might get bills from the PTT's and from the various hosts. With the diversity of data sources and communication routes, the potential for optimizing service and cost is great. This should lead to the creation of local distribution and accounting networks.

LOCAL DISTRIBUTION AND ACCOUNTING NETWORKS The local distribution networks could be public, commercial, or nonprofit. With the elements of accounting and electronic transfer of funds, the networks will probably develop as a new industry reflecting the diversity and orientation of the present sources of information for individuals. The library as the means of access to knowledge and information will undoubtedly assume new stature as the only place where this can be achieved at no cost to the individual. Since all libraries will be in a position of providing the same service within the same time frame, the only distinction may be the extent of the subsidy offered by the institution.

As a minimum, libraries should be prepared to offer access to the total index database in a manual microform retrieval mode. This minimum configuration would consist of the index at the rate of 200 or so frames per microfiche. Assuming that the initial database comprised 10/8 items, more than 20 times the present OCLC database, with indexing averaging 21 subject terms, or 10 proper names, and 7 geographic areas or segments, the total cost of the index database would be approximately $25,000 per library. This would be considerably less in the mass density store or even in high reduction microforms.

The index in library use, which would contain much more current material than the National Union Catalog, could be coupled with the microfiche interlibrary lending (MILL) system for large research libraries proposed by Edwin C. Pomranka (Proceedings 1978 of the Twenty-seventh Annual Conference of the National Micrographic Association, 43/33) to provide low cost access to library materials. One advantage of the MILL system is that the microfiche created to meet a specific request for interlibrary loan could be assigned the HUC identifier to be incorporated on each frame of the microfiche as a ready citation.

The microfiche could also be used as the input for a laser read/write mass density store of the type being developed for the Department of the Navy. Libraries could also provide access through terminals in the library connected to library networks offering facsimile transmission or documents in digital form.

Other local networks might be associated with newspapers, cable TV systems, television stations, new SDI enterprises, time share computer systems, research organizations, computer publishing companies, advertising agencies, etc..

In this type of configuration, the individual would make one contact whether by telephone or two way cable TV system. This one contact would provide access to the index and thence to any of the databases capable of supplying the desired information and without further telephone connection by the individual if data were to be provided by a telephone carrier. If it were to be a teletext delivery, as well might be the case for advertisements for goods or services, the page numbers would be automatically set in the terminal.

The local distribution network, which in most cases would handle all of the accounting, would have high speed access to the regional or system nodes for access to actual documents such as speeches, articles in other newspapers and periodicals, viewdata, and similar system information. It would have indexes to locally programamed teletext or FM subcarrier broadcast. The local network would handle all system queries either on line or by providing disks or other media for stand alone systems.

In serving as a complete information interface, the local distribution network could provide SDI service to its subscribers. It would provide a given amount of coverage in a daily dump in the users' systems at night with flash communication for urgent information. The daily dump would cover specific and general areas in proportions defined by the subscribers. Most items would have references for the complete text of abstracted or reviewed material, a speech summarized in a newspaper article, or a statistical table from which certain data had been extracted. By a short command, the user could order the material for immediate or later viewing.

If the subscriber had only a key pad without memory, the access to the total system would be provided frame by frame, as is the present case with viewdata systems. The assumption is that just as individuals wanted TV's in addition to radios, they will acquire personal computers and terminals with intelligent keyboards, and even stand alone systems complete with peripherals in addition to standard TV sets.

The local distribution and accounting systems will probably become important as input systems for companies desiring to sell goods and services. This is the area which will encourage the interplay between free information on teletext and FM subcarrier systems and two way digital systems where subscribers can place

orders, have goods delivered, and have payment made electronically with far more detailed records than one has today. For packaged goods and standard items, the user would not only save the trip to the supermarket and the department store, but would also have the potential, by calling special programs, of organizing menus and developing optimal food and other expenditure patterns for better nutrition and pleasure. The economy would save on fuel and pollution. New jobs would be created to replace those eliminated with the shuffling of boxes and packages and checking prices.

The special field of the HUC identification will accommodate the uniform product codes in use in the U.S. and being developed in Europe, so that each item ordered could have the identification of the person placing the order, the account to which payment is to be made, and the identification of the item. The computer could, of course, replace all code designations by full names and addresses, and provide electronic recapitulation of all expenditures.

STAND ALONE SYSTEMS, TERMINALS, AND INTELLIGENT KEYBOARDS As discussed above, the HUC system can greatly increase the capability of the individual to identify material to meet specific needs whether this is performed in the library at no cost to the individual or in the home at a variable and fixed cost depending upon the type of equipment and service desired.

Access to broadcast videotex, or teletext as it sometimes identified, and FM subcarrier broadcasts may well have only the fixed cost of equipment whether that is a modified TV set or a keyboard with other uses. Although access to both the index and substantive information can be obtained with a key pad, a keyboard with memory at a higher fixed cost would produce better results at a lower variable cost for the same access and would greatly extend the capability of the system for communication and other purposes.

In view of the diversification in types and structure of personal computers, the local distribution networks may specialize in serving a clientel owning particular systems. This would facilitate the identification and delivery of special software from the international database. Such distribution networks may actually develop from the user groups organized for different computer systems. Here again the HUC system offers the creator the opportunity to enter programs into the system and to be paid a royalty for use and reproduction. With the Cryptext device in every HUC terminal, the software can be easily programmed to run on an individual terminal. Since this would eliminate the potential for copying, software sold in this manner could be distributed to individuals by an online system such as MicroNET at a lower cost for the mass individual market.

Since many people will want certain information in hard copy to read while traveling or to use for handy reference as a time table or list of telephone numbers, many terminals in the home

will probably have a hard copy printer as a peripheral. For more extended copy, high speed on demand reproduction should be available.

IX. BENEFITS AND APPLICATIONS

As mentioned before, the effectiveness of any system for the user is determined by the cost and benefit. Cost is composed of two elements: fixed for the cost of hardware and software and variable for the cost of use. Benefit is measured by the ease of operation and effectiveness in producing the desired results. The desired results from the terminal in the home and at the place of work can range from providing entertainment to compiling sensor data in real time and producing immediate answers to complex questions. Since the HUC system is primarily concerned with the information process of collecting, storing, retrieving, communicating, and accounting as utilized by individuals untrained in computer science, the important elements in discussing benefit are the ease and operation of the system in guiding and assisting the user in performing these tasks and the effectiveness of the system in producing the desired result.

Assume a library terminal available to any user as is a card catalog today. This should be a stand alone unit which would automatically connect to a regional or system node for any query beyond its capacity. A single sign over the terminal stating "Press the HUC key and the ? key for assistance" would be sufficient for a new user to get started on a tree structure manual for the system. As a matter of fact, any entry on the keyboard that did not correspond with anything in the file would automatically produce an option list. The query system would be so designed that a neophyte could proceed step by step, while a sophisticated user or one with a specific query could proceed directly, to a specific section for reference aid, or with the process of retrieval or communication.

At any point, however, the user could request information on the meaning of any entries in the various fields of the code or the retrieval of desired items. Here cost and benefit are evident. If the termianl is a keyboard connected to a TV, then the amount of data which can be displayed is limited. This might require the transfer to the buffer of some or all of the material originally displayed. With a more expensive high resolution CRT, the double size characters could be reduced and all information displayed on the same screen.

One way of evaluating the effectiveness of the HUC system for the user is to examine the present system and compare it with the HUC concept for the following principal functions of a home information and computer center:

1. Selection of material to satisfy a specific need,

2. Accessing material or obtaining copies of documents,

3. Obtaining goods and services,

4. Using the system for specific tasks, and

5. Communicating with others.

SELECTION OF MATERIAL Requirements: Examples of user needs in selecting material include specific data known to exist such as telephone numbers, weather and market information, search for known material but not sufficiently identified for retrieval, search for unknown material to meet a specific requirement, browsing, entertainment, and current information.

Present: The present system for meeting these needs is characterized by the diversity of sources required, the lack of a unified index and the nonexistence of any index for a substantial portion of material which makes retrieval of known material in some areas almost impossible. Search for unknown material in certain computer databases is effective, but browsing is difficult. Selection of entertainment and news limited to the daily offerings of the industry. Viewdata and teletext offer promise but are limited by dependence on tree structure and lack of complete index.

HUC: This concept is characterized by complete index and immediate access to available databases, combines tree structure with direct alphabetical and digital access, provides for entertainment selection from the complete library of available cinema and video productions. All of this should be available in the home, library or at the place of work.

ACCESSING MATERIAL Requirements: To receive known data immediately at the least cost and documents, current information, and entertainment as a function of time, medium, and cost.

Present: The weakest aspect of the present system is the delivery of documents and entertainment. Books, even reference works, are often missing from libraries. Computer databases seldom provide complete texts of documents. Video and cinema productions are seldom available on demand. Viewdata and online systems providing the possibility of ordering documents is promising.

HUC: The on demand reproducing library and distribution centers supported by the mass density store should solve the document delivery problem. Cable TV with the possibility of dumping cinema and video productions in specific terminals should provide a solution to entertainment delivery.

OBTAINING GOODS AND SERVICES Requirements: Delivery of packaged food and other essentials, locating bargains, obtaining professional advice, locating required expertise.

Present: Little food delivery, much catalog ordering at high distribution cost for advertising, yellow pages, books and letters for professional advice, viewdata, teletext, and Technotec.

HUC: Low advertising cost for input into system, electronic funds transfer with complete accounting, closed user group for legal services, individual input similar to classified advertising, biographic, and resume information.

USING THE SYSTEM Requirements: Supervising household functions, optimizing expenditure patterns, controlling funds, and generating income.

Present: Just beginning with home computer systems, concept of telesoftware, Digital Broadcast project, electronic banking, computer user groups, and word processing.

HUC: The problem of using home computer systems for these functions is a combination of software and systems support. By providing the index to software and a system in which creators of software and other intellectual works can receive compensation, the HUC concept contributes to the more effective use of home and other computer systems.

COMMUNICATION Requirements: Quick, secure, inexpensive, and records.

Present: Telex, neither secure nor inexpensive, inter-computer communication excellent and inexpensive but nothing for individuals.

HUC: This concept offers addressing by individual, organization or terminal with encryption of communication and positive terminal identification. Correspondence records and control is facilitated, and the individual is in a position to enjoy the quick, inexpensive, and secure service that is almost the same as enjoyed today by large computer systems.

X. CONCLUSIONS

It is evident that we are on the verge of a new wave of social change which will emerge from the technical developments of the past decade. It will be similar to the change that took place with the advent of television when some ten thousand TV sets suddenly became ten million with all of the infrastructure of new TV stations. The mainspring of this innovative development will be the videotext with its development in Europe, Canada, and Japan. For the development in the United States to penetrate the mass market, it will be necessary to provide the hardware and software designed to meet the requirements of a literate society.

Action which could be taken now to produce results within a one year time frame are summarized as follows:

1. Development of the HUC software;
2. Organization of a portion of the database;
3. Development of prototypes of the terminals;
4. Test of the concept on developing systems;
5. Incorporation in economic development projects;
6. Experimentation with various mass density storage systems; and
7. Test of communication concept with packet switching and correspondence control.